

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: TAKAHASHI, Tohru  
SERIZAWA, Nobufusa  
KOISHI, Ryuta  
KAWASHIMA, Ichiro
- (ii) TITLE OF INVENTION: EXPRESSION SYSTEMS UTILIZING  
AUTOLYZING FUSION PROTEINS  
AND A NOVEL REDUCING POLYPEPTIDE

(iii) NUMBER OF SEQUENCES: 19

- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Frishhauf, Holtz, Goodman, Langer & Chick, P.C.
  - (B) STREET: 767 Third Avenue-25th Floor
  - (C) CITY: New York
  - (D) STATE: New York
  - (E) COUNTRY: United States
  - (F) ZIP: 10017-2023
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.24
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/500,635
  - (B) FILING DATE: 11-JUL-1995
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: JP 6-161053
  - (B) FILING DATE: 13-JUL-1994
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: JP 6-218392
  - (B) FILING DATE: 13-SEP-1994
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: JP 6-303809
  - (B) FILING DATE: 07-DEC-1994
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Goodman, Herbert
  - (B) REGISTRATION NUMBER: 17081
  - (C) REFERENCE/DOCKET NUMBER: 950376/HG
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (212) 319-4900
  - (B) TELEFAX: (212) 319-5101
  - (C) TELEX: 236268

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1320 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Clover Yellow Vein Virus

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1320

(D) OTHER INFORMATION:

(ix) FEATURE:

(A) NAME/KEY: mat\_peptide

(B) LOCATION: 10..1311

(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAG	TTC	CAA	GGG	AAA	AGT	AAG	AGA	ACA	AGA	CAA	AAG	TTG	AAG	TTC	AGA	48
Lys	Phe	Gln	Gly	Lys	Ser	Lys	Arg	Thr	Arg	Gln	Lys	Leu	Lys	Phe	Arg	
1				5				10						15		
GCG	GCA	AGA	GAC	ATG	AAG	GAT	CGT	TAT	GAA	GTG	CAT	GCC	GAT	GAG	GGG	96
Ala	Ala	Arg	Asp	Met	Lys	Asp	Arg	Tyr	Glu	Val	His	Ala	Asp	Glu	Gly	
				20				25					30			
ACT	TTA	GTG	GAA	AAT	TTT	GGA	ACT	CGT	TAT	TCA	AAG	AAA	GGC	AAG	ACA	144
Thr	Leu	Val	Glu	Asn	Phe	Gly	Thr	Arg	Tyr	Ser	Lys	Lys	Gly	Lys	Thr	
				35				40				45				
AAA	GGT	ACT	GTT	GTG	GGT	TTG	GGT	GCA	AAA	ACA	AGA	CGG	TTC	ACT	AAC	192
Lys	Gly	Thr	Val	Val	Gly	Leu	Gly	Ala	Lys	Thr	Arg	Arg	Phe	Thr	Asn	
				50			55				60					
ATG	TAT	GGT	TTT	GAC	CCC	ACG	GAG	TAT	TCA	TTT	GCT	AGG	TAT	CTT	GAT	240
Met	Tyr	Gly	Phe	Asp	Pro	Thr	Glu	Tyr	Ser	Phe	Ala	Arg	Tyr	Leu	Asp	
				65			70			75		80				
CCA	ATC	ACG	GGT	GCA	ACA	TTG	GAT	GAA	ACC	CCA	ATT	CAC	AAT	GTA	AAT	288
Pro	Ile	Thr	Gly	Ala	Thr	Leu	Asp	Glu	Thr	Pro	Ile	His	Asn	Val	Asn	
				85			90				95					
TTG	GTT	GCT	GAG	CAT	TTT	GGC	GAC	ATA	AGG	CTT	GAT	ATG	GTT	GAC	AAG	336
Leu	Val	Ala	Glu	His	Phe	Gly	Asp	Ile	Arg	Leu	Asp	Met	Val	Asp	Lys	
				100			105				110					
GAG	TTA	CTT	GAC	AAA	CAG	CAC	TTA	TAC	CTC	AAG	AGA	CCA	ATA	GAA	TGT	384
Glu	Leu	Leu	Asp	Lys	Gln	His	Leu	Tyr	Leu	Lys	Arg	Pro	Ile	Glu	Cys	
				115			120			125						
TAC	TTT	GTA	AAG	GAT	GCT	GGT	CAG	AAG	GTG	ATG	AGG	ATT	GAT	CTA	ACA	432
Tyr	Phe	Val	Lys	Asp	Ala	Gly	Gln	Lys	Val	Met	Arg	Ile	Asp	Leu	Thr	
				130			135			140						
CCC	CAC	AAC	CCA	TTG	TTG	GCA	AGC	GAT	GTT	AGC	ACA	ACC	ATA	ATG	GGT	480
Pro	His	Asn	Pro	Leu	Leu	Ala	Ser	Asp	Val	Ser	Thr	Thr	Ile	Met	Gly	
				145			150			155		160				

TAT CCT GAG AGA GAA GGT GAA CTC CGT CAA ACT GGA AAG GCA AGG TTA Tyr Pro Glu Arg Glu Gly Glu Leu Arg Gln Thr Gly Lys Ala Arg Leu 165 170 175	528
GTC GAC CCA TCA GAG TTG CCC GCG CGG AAT GAG GAT ATT GAT GCA GAG Val Asp Pro Ser Glu Leu Pro Ala Arg Asn Glu Asp Ile Asp Ala Glu 180 185 190	576
TTT GAG AGT CTA AAT CGC ATA AGT GGT TTG CGC GAC TAT AAT CCC ATT Phe Glu Ser Leu Asn Arg Ile Ser Gly Leu Arg Asp Tyr Asn Pro Ile 195 200 205	624
TCA CAA AAT GTT TGC TTG CTA ACA AAT GAG TCA GAA GGC CAT AGA GAG Ser Gln Asn Val Cys Leu Leu Thr Asn Glu Ser Gly His Arg Glu 210 215 220	672
AAG ATG TTT GGA ATT GGA TAT GGT TCA GTG ATC ATT ACA AAT CAA CAT Lys Met Phe Gly Ile Gly Tyr Gly Ser Val Ile Ile Thr Asn Gln His 225 230 235 240	720
CTG TTC AGA AGG AAT AAT GGG GAG TTA TCA ATT CAA TCC AAG CAT GGC Leu Phe Arg Arg Asn Asn Gly Glu Leu Ser Ile Gln Ser Lys His Gly 245 250 255	768
TAC TTC AGA TGC CGC AAC ACC ACA AGC TTG AAG ATG CTG CCT TTG GAG Tyr Phe Arg Cys Arg Asn Thr Ser Leu Lys Met Leu Pro Leu Glu 260 265 270	816
GGA CAT GAC ATT TTG TTG ATT CAG TTA CCA AGG GAC TTT CCA GTG TTT Gly His Asp Ile Leu Leu Ile Gln Leu Pro Arg Asp Phe Pro Val Phe 275 280 285	864
CCA CAA AAG ATT CGC TTT AGG GAG CCA AGA GTG GAT GAC AAA ATT GTT Pro Gln Lys Ile Arg Phe Arg Glu Pro Arg Val Asp Asp Lys Ile Val 290 295 300	912
TTG GTC AGC ACA AAT TTC CAG GAA AAG AGT TCC TCG AGC ACG GTC TCA Leu Val Ser Thr Asn Phe Gln Glu Lys Ser Ser Ser Thr Val Ser 305 310 315 320	960
GAG TCC AGT AAC ATT TCA AGA GTG CAG TCA GCC AAT TTC TAC AAG CAT Glu Ser Ser Asn Ile Ser Arg Val Gln Ser Ala Asn Phe Tyr Lys His 325 330 335	1008
TGG ATC TCA ACA GTA GCA GGA CAC TGT GGA AAC CCT ATG GTT TCG ACT Trp Ile Ser Thr Val Ala Gly His Cys Gly Asn Pro Met Val Ser Thr 340 345 350	1056
AAA GAT GGA TTT ATT GTA GGT ATC CAC AGT CTT GCT TCA TTG ACA GGC Lys Asp Gly Phe Ile Val Gly Ile His Ser Leu Ala Ser Leu Thr Gly 355 360 365	1104
GAC GTT AAC ATC TTC ACA AGC TTT CCG CCG CAG TTT GAG AAC AAA TAT Asp Val Asn Ile Phe Thr Ser Phe Pro Pro Gln Phe Glu Asn Lys Tyr 370 375 380	1152
CTA CAG AAG CTC AGT GAA CAC ACA TGG TGT AGT GGA TGG AAA CTA AAT Leu Gln Lys Leu Ser Glu His Thr Trp Cys Ser Gly Trp Lys Leu Asn 385 390 395 400	1200
CTT GGA AAG ATT AGT TGG GGT GGA ATC AAC ATT GTG GAG GAT GCA CCT Leu Gly Lys Ile Ser Trp Gly Gly Ile Asn Ile Val Glu Asp Ala Pro 405 410 415	1248

GAA GAG CCC TTT ATA ACA TCC AAG ATG GCA AGC CTT CTT AGT GAT TTG 1296  
Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu  
420 425 430

AAT TGT TCA TTC CAA GCA AGT GCG 1320  
Asn Cys Ser Phe Gln Ala Ser Ala  
435 440

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 440 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Clover Yellow Vein Virus

(ix) FEATURE:  
(A) NAME/KEY: mat\_peptide  
(B) LOCATION: 4..437  
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys Phe Gln Gly Lys Ser Lys Arg Thr Arg Gln Lys Leu Lys Phe Arg  
1 5 10 15

Ala Ala Arg Asp Met Lys Asp Arg Tyr Glu Val His Ala Asp Glu Gly  
20 25 30

Thr Leu Val Glu Asn Phe Gly Thr Arg Tyr Ser Lys Lys Gly Lys Thr  
35 40 45

Lys Gly Thr Val Val Gly Leu Gly Ala Lys Thr Arg Arg Phe Thr Asn  
50 55 60

Met Tyr Gly Phe Asp Pro Thr Glu Tyr Ser Phe Ala Arg Tyr Leu Asp  
65 70 75 80

Pro Ile Thr Gly Ala Thr Leu Asp Glu Thr Pro Ile His Asn Val Asn  
85 90 95

Leu Val Ala Glu His Phe Gly Asp Ile Arg Leu Asp Met Val Asp Lys  
100 105 110

Glu Leu Leu Asp Lys Gln His Leu Tyr Leu Lys Arg Pro Ile Glu Cys  
115 120 125

Tyr Phe Val Lys Asp Ala Gly Gln Lys Val Met Arg Ile Asp Leu Thr  
130 135 140

Pro His Asn Pro Leu Leu Ala Ser Asp Val Ser Thr Thr Ile Met Gly  
145 150 155 160

Tyr Pro Glu Arg Glu Gly Glu Leu Arg Gln Thr Gly Lys Ala Arg Leu  
165 170 175

Val Asp Pro Ser Glu Leu Pro Ala Arg Asn Glu Asp Ile Asp Ala Glu  
180 185 190

Phe Glu Ser Leu Asn Arg Ile Ser Gly Leu Arg Asp Tyr Asn Pro Ile  
195 200 205  
Ser Gln Asn Val Cys Leu Leu Thr Asn Glu Ser Glu Gly His Arg Glu  
210 215 220  
Lys Met Phe Gly Ile Gly Tyr Gly Ser Val Ile Ile Thr Asn Gln His  
225 230 235 240  
Leu Phe Arg Arg Asn Asn Gly Glu Leu Ser Ile Gln Ser Lys His Gly  
245 250 255  
Tyr Phe Arg Cys Arg Asn Thr Thr Ser Leu Lys Met Leu Pro Leu Glu  
260 265 270  
Gly His Asp Ile Leu Leu Ile Gln Leu Pro Arg Asp Phe Pro Val Phe  
275 280 285  
Pro Gln Lys Ile Arg Phe Arg Glu Pro Arg Val Asp Asp Lys Ile Val  
290 295 300  
Leu Val Ser Thr Asn Phe Gln Glu Lys Ser Ser Ser Ser Thr Val Ser  
305 310 315 320  
Glu Ser Ser Asn Ile Ser Arg Val Gln Ser Ala Asn Phe Tyr Lys His  
325 330 335  
Trp Ile Ser Thr Val Ala Gly His Cys Gly Asn Pro Met Val Ser Thr  
340 345 350  
Lys Asp Gly Phe Ile Val Gly Ile His Ser Leu Ala Ser Leu Thr Gly  
355 360 365  
Asp Val Asn Ile Phe Thr Ser Phe Pro Pro Gln Phe Glu Asn Lys Tyr  
370 375 380  
Leu Gln Lys Leu Ser Glu His Thr Trp Cys Ser Gly Trp Lys Leu Asn  
385 390 395 400  
Leu Gly Lys Ile Ser Trp Gly Gly Ile Asn Ile Val Glu Asp Ala Pro  
405 410 415  
Glu Glu Pro Phe Ile Thr Ser Lys Met Ala Ser Leu Leu Ser Asp Leu  
420 425 430  
Asn Cys Ser Phe Gln Ala Ser Ala  
435 440

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTCCATGGGG AAAAGTAAGA GAACA

25

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ACTCTGAGAC CGTGCTCGAG

20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGGAAAAGAG TTCCTCGAGC

20

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AATTGTTCAT TCCAAGCACC TGGGCCACCA CCTGGC

36

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCCAGGTGGT GGCCCAGGTG CTTGGAATGA ACAATT

36

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTGTCAGCAC ACCTGGGAGC TGTAGAGCTC

30

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ala Pro Gly Pro Pro Pro Gly  
1 5

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Pro Gly Pro Pro Pro Gly Pro  
1 5

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1650 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens
- (H) CELL LINE: KM-102

(vii) IMMEDIATE SOURCE:  
(B) CLONE: KM31-7

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1647
- (D) OTHER INFORMATION:

(ix) FEATURE:

- (A) NAME/KEY: mat\_peptide
- (B) LOCATION: 70..1647
- (D) OTHER INFORMATION:

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 1..69
- (D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATG TCA TGT GAG GAC GGT CGG GCC CTG GAA GGA ACG CTC TCG GAA TTG  
Met Ser Cys Glu Asp Gly Arg Ala Leu Glu Gly Thr Leu Ser Glu Leu  
-23 -20 -15 -10

48

GCC GCG GAA ACC GAT CTG CCC GTT GTG TTT GTG AAA CAG AGA AAG ATA  
Ala Ala Glu Thr Asp Leu Pro Val Val Phe Val Lys Gln Arg Lys Ile  
-5 1 5

96

GGC GGC CAT GGT CCA ACC TTG AAG GCT TAT CAG GAG GGC AGA CTT CAA  
Gly Gly His Gly Pro Thr Leu Lys Ala Tyr Gln Glu Gly Arg Leu Gln  
10 15 20 25

144

DRAFT - OFFICIAL

AAG CTA CTA AAA ATG AAC GGC CCT GAA GAT CTT CCC AAG TCC TAT GAC Lys Leu Leu Lys Met Asn Gly Pro Glu Asp Leu Pro Lys Ser Tyr Asp 30 35 40	192
TAT GAC CTT ATC ATC ATT GGA GGT GGC TCA GGA GGT CTG GCA GCT GCT Tyr Asp Leu Ile Ile Ile Gly Gly Ser Gly Gly Leu Ala Ala Ala 45 50 55	240
AAG GAG GCA GCC CAA TAT GGC AAG AAG GTG ATG GTC CTG GAC TTT GTC Lys Glu Ala Ala Gln Tyr Gly Lys Lys Val Met Val Leu Asp Phe Val 60 65 70	288
ACT CCC ACC CCT CTT GGA ACT AGA TGG GGT CTT GGA GGA ACA TGT GTG Thr Pro Thr Pro Leu Gly Thr Arg Trp Gly Leu Gly Gly Thr Cys Val 75 80 85	336
AAT GTG GGT TGC ATA CCT AAA AAA CTG ATG CAT CAA GCA GCT TTG TTA Asn Val Gly Cys Ile Pro Lys Lys Leu Met His Gln Ala Ala Leu Leu 90 95 100 105	384
GGA CAA GCC CTG CAA GAC TCT CGA AAT TAT GGA TGG AAA GTC GAG GAG Gly Gln Ala Leu Gln Asp Ser Arg Asn Tyr Gly Trp Lys Val Glu Glu 110 115 120	432
ACA GTT AAG CAT GAT TGG GAC AGA ATG ATA GAA GCT GTA CAG AAT CAC Thr Val Lys His Asp Trp Asp Arg Met Ile Glu Ala Val Gln Asn His 125 130 135	480
ATT GGC TCT TTG AAT TGG GGC TAC CGA GTA GCT CTG CGG GAG AAA AAA Ile Gly Ser Leu Asn Trp Gly Tyr Arg Val Ala Leu Arg Glu Lys Lys 140 145 150	528
GTC GTC TAT GAG AAT GCT TAT GGG CAA TTT ATT GGT CCT CAC AGG ATT Val Val Tyr Glu Asn Ala Tyr Gly Gln Phe Ile Gly Pro His Arg Ile 155 160 165	576
AAG GCA ACA AAT AAT AAA GGC AAA GAA AAA ATT TAT TCA GCA GAG AGA Lys Ala Thr Asn Asn Lys Gly Lys Glu Lys Ile Tyr Ser Ala Glu Arg 170 175 180 185	624
TTT CTC ATT GCC ACT GGT GAA AGA CCA CGT TAC TTG GGC ATC CCT GGT Phe Leu Ile Ala Thr Gly Glu Arg Pro Arg Tyr Leu Gly Ile Pro Gly 190 195 200	672
GAC AAA GAA TAC TGC ATC AGC AGT GAT GAT CTT TTC TCC TTG CCT TAC Asp Lys Glu Tyr Cys Ile Ser Ser Asp Asp Leu Phe Ser Leu Pro Tyr 205 210 215	720
TGC CCG GGT AAG ACC CTG GTT GGA GCA TCC TAT GTC GCT TTG GAG Cys Pro Gly Lys Thr Leu Val Val Gly Ala Ser Tyr Val Ala Leu Glu 220 225 230	768
TGC GCT GGA TTT CTT GCT GGT ATT GGT TTA GAC GTC ACT GTT ATG GTT Cys Ala Gly Phe Leu Ala Gly Ile Gly Leu Asp Val Thr Val Met Val 235 240 245	816
AGG TCC ATT CTT CTT AGA GGA TTT GAC CAG GAC ATG GCC AAC AAA ATT Arg Ser Ile Leu Leu Arg Gly Phe Asp Gln Asp Met Ala Asn Lys Ile 250 255 260 265	864
GGT GAA CAC ATG GAA CAT GGC ATC AAG TTT ATA AGA CAG TTC GTA Gly Glu His Met Glu Glu His Gly Ile Lys Phe Ile Arg Gln Phe Val 270 275 280	912

CCA ATT AAA GTT GAA CAA ATT GAA GCA GGG ACA CCA GGC CGA CTC AGA Pro Ile Lys Val Glu Gln Ile Glu Ala Gly Thr Pro Gly Arg Leu Arg 285	290	295	960	
GTA GTA GCT CAG TCC ACC AAT AGT GAG GAA ATC ATT GAA GGA GAA TAT Val Val Ala Gln Ser Thr Asn Ser Glu Glu Ile Ile Glu Gly Glu Tyr 300	305	310	1008	
AAT ACG GTG ATG CTG GCA ATA GGA AGA GAT GCT TGC ACA AGA AAA ATT Asn Thr Val Met Leu Ala Ile Gly Arg Asp Ala Cys Thr Arg Lys Ile 315	320	325	1056	
GGC TTA GAA ACC GTA GGG GTG AAG ATA AAT GAA AAG ACT GGA AAA ATA Gly Leu Glu Thr Val Gly Val Lys Ile Asn Glu Lys Thr Gly Lys Ile 330	335	340	345	1104
CCT GTC ACA GAT GAA GAA CAG ACC AAT GTG CCT TAC ATC TAT GCC ATT Pro Val Thr Asp Glu Glu Gln Thr Asn Val Pro Tyr Ile Tyr Ala Ile 350	355	360	1152	
GGC GAT ATA TTG GAG GAT AAG GTG GAG CTC ACC CCA GTT GCA ATC CAG Gly Asp Ile Leu Glu Asp Lys Val Glu Leu Thr Pro Val Ala Ile Gln 365	370	375	1200	
GCA GGA AGA TTG CTG GCT CAG AGG CTC TAT GCA GGT TCC ACT GTC AAG Ala Gly Arg Leu Leu Ala Gln Arg Leu Tyr Ala Gly Ser Thr Val Lys 380	385	390	1248	
TGT GAC TAT GAA AAT GTT CCA ACC ACT GTA TTT ACT CCT TTG GAA TAT Cys Asp Tyr Glu Asn Val Pro Thr Thr Val Phe Thr Pro Leu Glu Tyr 395	400	405	1296	
GGT GCT TGT GGC CTT TCT GAG GAG AAA GCT GTG GAG AAG TTT GGG GAA Gly Ala Cys Gly Leu Ser Glu Glu Lys Ala Val Glu Lys Phe Gly Glu 410	415	420	425	1344
GAA AAT ATT GAG GTT TAC CAT AGT TAC TTT TGG CCA TTG GAA TGG ACG Glu Asn Ile Glu Val Tyr His Ser Tyr Phe Trp Pro Leu Glu Trp Thr 430	435	440	1392	
ATT CCG TCA AGA GAT AAC AAC AAA TGT TAT GCA AAA ATA ATC TGT AAT Ile Pro Ser Arg Asp Asn Asn Lys Cys Tyr Ala Lys Ile Ile Cys Asn 445	450	455	1440	
ACT AAA GAC AAT GAA CGT GTT GTG GGC TTT CAC GTA CTG GGT CCA AAT Thr Lys Asp Asn Glu Arg Val Val Gly Phe His Val Leu Gly Pro Asn 460	465	470	1488	
GCT GGA GAA GTT ACA CAA GGC TTT GCA GCT GCG CTC AAA TGT GGA CTG Ala Gly Glu Val Thr Gln Gly Phe Ala Ala Leu Lys Cys Gly Leu 475	480	485	1536	
ACC AAA AAG CAG CTG GAC AGC ACA ATT GGA ATC CAC CCT GTC TGT GCA Thr Lys Lys Gln Leu Asp Ser Thr Ile Gly Ile His Pro Val Cys Ala 490	495	500	505	1584
GAG GTA TTC ACA ACA TTG TCT GTG ACC AAG CGC TCT GGG GCA AGC ATC Glu Val Phe Thr Thr Leu Ser Val Thr Lys Arg Ser Gly Ala Ser Ile 510	515	520	1632	
CTC CAG GCT GGC TGC TGA Leu Gln Ala Gly Cys 525			1650	

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Ser Cys Glu Asp Gly Arg Ala Leu Glu Gly Thr Leu Ser Glu Leu  
-23            -20            -15            -10

Ala Ala Glu Thr Asp Leu Pro Val Val Phe Val Lys Gln Arg Lys Ile  
-5              1              5

Gly Gly His Gly Pro Thr Leu Lys Ala Tyr Gln Glu Gly Arg Leu Gln  
10            15            20            25

Lys Leu Leu Lys Met Asn Gly Pro Glu Asp Leu Pro Lys Ser Tyr Asp  
30            35            40

Tyr Asp Leu Ile Ile Gly Gly Ser Gly Gly Leu Ala Ala Ala  
45            50            55

Lys Glu Ala Ala Gln Tyr Gly Lys Lys Val Met Val Leu Asp Phe Val  
60            65            70

Thr Pro Thr Pro Leu Gly Thr Arg Trp Gly Leu Gly Gly Thr Cys Val  
75            80            85

Asn Val Gly Cys Ile Pro Lys Lys Leu Met His Gln Ala Ala Leu Leu  
90            95            100            105

Gly Gln Ala Leu Gln Asp Ser Arg Asn Tyr Gly Trp Lys Val Glu Glu  
110            115            120

Thr Val Lys His Asp Trp Asp Arg Met Ile Glu Ala Val Gln Asn His  
125            130            135

Ile Gly Ser Leu Asn Trp Gly Tyr Arg Val Ala Leu Arg Glu Lys Lys  
140            145            150

Val Val Tyr Glu Asn Ala Tyr Gly Gln Phe Ile Gly Pro His Arg Ile  
155            160            165

Lys Ala Thr Asn Asn Lys Gly Lys Glu Lys Ile Tyr Ser Ala Glu Arg  
170            175            180            185

Phe Leu Ile Ala Thr Gly Glu Arg Pro Arg Tyr Leu Gly Ile Pro Gly  
190            195            200

Asp Lys Glu Tyr Cys Ile Ser Ser Asp Asp Leu Phe Ser Leu Pro Tyr  
205            210            215

Cys Pro Gly Lys Thr Leu Val Val Gly Ala Ser Tyr Val Ala Leu Glu  
220            225            230

Cys Ala Gly Phe Leu Ala Gly Ile Gly Leu Asp Val Thr Val Met Val  
235            240            245

Arg Ser Ile Leu Leu Arg Gly Phe Asp Gln Asp Met Ala Asn Lys Ile  
 250 255 260 265  
 Gly Glu His Met Glu Glu His Gly Ile Lys Phe Ile Arg Gln Phe Val  
 270 275 280  
 Pro Ile Lys Val Glu Gln Ile Glu Ala Gly Thr Pro Gly Arg Leu Arg  
 285 290 295  
 Val Val Ala Gln Ser Thr Asn Ser Glu Glu Ile Ile Glu Gly Glu Tyr  
 300 305 310  
 Asn Thr Val Met Leu Ala Ile Gly Arg Asp Ala Cys Thr Arg Lys Ile  
 315 320 325  
 Gly Leu Glu Thr Val Gly Val Lys Ile Asn Glu Lys Thr Gly Lys Ile  
 330 335 340 345  
 Pro Val Thr Asp Glu Glu Gln Thr Asn Val Pro Tyr Ile Tyr Ala Ile  
 350 355 360  
 Gly Asp Ile Leu Glu Asp Lys Val Glu Leu Thr Pro Val Ala Ile Gln  
 365 370 375  
 Ala Gly Arg Leu Leu Ala Gln Arg Leu Tyr Ala Gly Ser Thr Val Lys  
 380 385 390  
 Cys Asp Tyr Glu Asn Val Pro Thr Thr Val Phe Thr Pro Leu Glu Tyr  
 395 400 405  
 Gly Ala Cys Gly Leu Ser Glu Glu Lys Ala Val Glu Lys Phe Gly Glu  
 410 415 420 425  
 Glu Asn Ile Glu Val Tyr His Ser Tyr Phe Trp Pro Leu Glu Trp Thr  
 430 435 440  
 Ile Pro Ser Arg Asp Asn Asn Lys Cys Tyr Ala Lys Ile Ile Cys Asn  
 445 450 455  
 Thr Lys Asp Asn Glu Arg Val Val Gly Phe His Val Leu Gly Pro Asn  
 460 465 470  
 Ala Gly Glu Val Thr Gln Gly Phe Ala Ala Leu Lys Cys Gly Leu  
 475 480 485  
 Thr Lys Lys Gln Leu Asp Ser Thr Ile Gly Ile His Pro Val Cys Ala  
 490 495 500 505  
 Glu Val Phe Thr Thr Leu Ser Val Thr Lys Arg Ser Gly Ala Ser Ile  
 510 515 520  
 Leu Gln Ala Gly Cys  
 525

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TAAATAAAATA AATAA

15

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTAGCGCTCT GGGGCAAGCA TCCTCCAGGC TGGCTGCCAC CACCACCAACC ACCACTGATC

60

TAGACT

66

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGTCAGCACA AATTTCGA

18

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AAACACAAC TGGAAATGAAC AATT

24

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TCATTCCAAG TTGTGTTTGT GAAA

24

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, synthetic DNA

(iii) HYPOTHETICAL: N

(iv) ANTI-SENSE: N

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CATAGGATGC TCCAACAA

18

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Asn Cys Ser Phe Gln Xaa  
1 5